ALTERNATÍVNE METÓDY URČOVANIA OBSAHU VODY V PÔDE

Martin Suchár

Predpovedanie obsahu vody v pôde má zásadný význam pre vývoj pokročilých poľnohospodárskych informačných systémov a pre modelovanie v hydrológii. Pretože obsah vody v pôde je premenlivý a nestacionárny, je pri jeho predpovedi ťažké dosiahnuť presné výsledky. Vzhľadom k tomu, sú stále potrebné nové a presnejšie modelovacie metódy, ktoré spresnia výslednú predpoveď. Autori sa v príspevku venujú určovaniu pedotransférnych funkcií, ktoré umožňujú stanoviť hydraulické vlastnosti pôdy z ľahšie merateľných pôdnych vlastností a sú dôležitým prvkom pre uvedené modelovanie. V nedávnej dobe sa objavili snahy využívať pre ich určenie umelé neurónové siete (UNS). Novší vývoj v oblasti učiacich algoritmov poskytuje ďalšie možnosti, z ktorých sa v tomto príspevku venujeme tzv. mechanizmom podporných vektorov (Support Vector Machines – SVM). Tieto sa najskôr používali ako klasifikátor a neskôr sa ich funkčnosť rozšírila aj na regresiu. Presnosť SVM budeme porovnávať s viacnásobnou lineárnou regresiou.

KLÚČOVÉ SLOVÁ: vlhkosť pôdy, pedotransférna krivka, predpovedanie, support vector machines

ALTERNATIVE METHODS FOR DETERMINING WATER CONTENT IN SOIL

Predicting water content in soil is essential for the development of advanced agricultural information systems and modeling in hydrology. Because the water content in soil is variable and non-stationary, it is difficult to achieve accurate results in the forecast. Due to these problems are still necessary new and better learning methods to specify the resulting forecast. Pedotransfer functions (PTFs), which estimate soil hydraulic parameters from easy to measure soil properties, are an important data source for hydrologic modeling. Recently artificial neural networks (ANNs) have become the tool of choice in PTF development. Recent developments in machine learning methods include the growing research and application of the alternative data-driven method called Support Vector Machines (SVMs). SVMs were first used as a classifier and then their functionality has been extended to regression. We compare the accuracy of the SVM with older learner technology, multi-layer perceptron and also with multiple linear regression.

KEY WORDS: soil moisture, pedotransfer function, forecasting, support vector machine

Úvod

Hydrologické procesy, sú vo všeobecnosti nelineárne a riadia sa všeobecnými fyzikálnymi zákonitosťami. Ich variabilita pokrýva viacero časových a priestorových úrovní. Preto je náročné ich pozorovať, charakterizovať a modelovať. V súčasnosti je jeden z najprijateľnejších spôsobov skúmania takýchto procesov na rozľahlých územiach možný pomocou modelovania. Aby sa vytvorené modely čo najviac približovali skutočnosti je nutné do nich zadať všetky potrebné vstupné údaje, ktoré ich charakterizujú. Analyzovanie takéhoto - často veľkého množstva dát predstavuje značný problém, takže nároky na presnejšie a výkonné algoritmy sú naozaj opodstatnené, keďže výsledná presnosť predpovede nezávisí len od vstupných údajov, ale aj od výberu vhodného výpočtového modelu. Výhodou modelovania pomocou umelých neurónových sietí (UNS) je ich schopnosť učiť sa zo vzorových údajov a tiež schopnosť poznatky z nich zovšeobecniť, často bez detailného poznania rôznych stavových veličín o procese, čo z nich robí vhodný alternatívny nástroj na riešenie zložitých procesov (Šimor a kol., 2009).

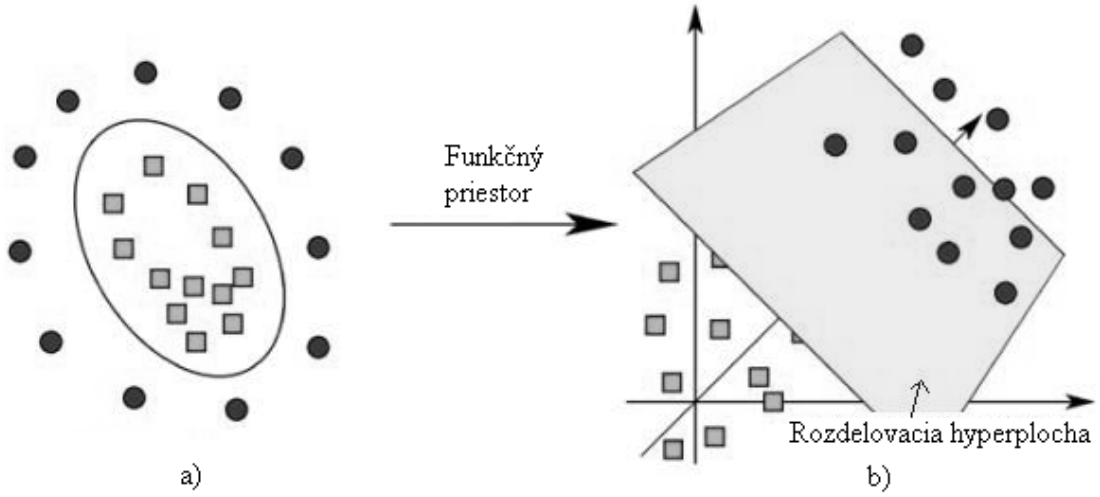
Informácie o obsahu pôdnej vody sú pri plánovaní využívania vodných zdrojov a hospodárenia s nimi veľmi cenné. Modelovanie a predpovedanie prenosu pôdnej vody je veľmi dôležité v poľnohospodárstve, či hydrológii. V poľnohospodárstve môže ísť o plánovanie zavlažovacích systémov a predpovedanie výnosov pomocou presných predpovedí o budúcich stavoch pôdnej vody.

V hydrológii, sú tieto poznatky nevyhnutné pre analýzu zrážok, odtokového procesu či riadenie vodných zdrojov. Jednou z najdôležitejších hydrofyzikálnych charakteristík pôdy, ktorá vstupuje do hydraulických alebo hydrologických modelov je vlhkostná retenčná krivka (VRK), ktorá vyjadruje závislosť energie pôdnej vody od obsahu vody, teda vzťah medzi vlhkosťou pôdy a vlhkostným potenciálom. Pre potreby modelovania je vhodné vyjadriť jej priebeh analyticky (alebo algoritmicky). Závisí od štruktúry pôdy, jej objemovej hmotnosti, od zrnitostného a mineralogického zloženia, od obsahu humusu a od obsahu a vlastností katiónov. Priebeh retenčných čiar stanovujeme štandardne laboratórne. Metóda určenia bodov vlhkostnej retenčnej krivky v laboratórnych podmienkach je však náročná na experimentálne zariadenie, na čas potrebný pre použitie tejto metódy a je veľmi pracná. V pôdnej fyzike sa preto vypracovali metódy na určovanie hydrofyzikálnych charakteristík pôdy v závislosti od jej zrnitostného zloženia, obsahu organických látok a objemovej hmotnosti (Brooks, Corey, 1964; Renger, 1971; Gupta, Larson, 1979; Rawls a kol., 1982, Wosten a kol., 1998). Pre tieto modely (alebo vzťahy) sa zaviedol termín pedotransférne funkcie (PTF), t.j. funkcie, ktoré uvádzajú do vzťahu rôzne pôdne charakteristiky a vlastnosti jednu k druhej (Bouma aVan Lanen, 1987). PTF sú funkčnou závislosťou, ktorá prostredníctvom prístupných pôdnych vlastností (napr. textúra, štruktúra, organická hmota) umožňuje vyjadriť chýbajúce pôdne vlastnosti (napr. hydraulické a chemické vlastnosti). Podrobnejší popis je dostupný aj v slovenčine – napr. Skalová, J. (2001).
V predkladanej práci porovnávame vytváranie týchto funkčných závislostí pomocou support vector machines (SVM) a štandardnej multilineárnej regresie.
PTF pôd Záhorskej nížiny a ich aplikáciu pri modelovaní vodného režimu pôdy rozpracovala Skalová (2001) odkiaľ sme prevzali vstupné dáta a to percentá obsahu zrnitostných kategórií I až IV podľa Kopeckého, redukovanú objemovú hmotnosť (ρ_d) a vlhkosti pre vlkostné potenciály h_w= -2.5, -56, -209, -558, -976, -3060, -15300 cm stanovené laboratórne pre potreby určenia a testovania regresných závislostí.

Support vector machine

Support vector machine (SVM), čiže mechanizmy podporných vektorov, sú novším typom dátovo riadených modelov schopných učiť sa. Ich prvé využitie bolo zamerané na klasifikáciu a neskôr bolo rozšírené na regresiu. Čo do štruktúry sú prakticky zhodné s UNS, avšak majú iný spôsob učenia, ktorý orientačne v ďalšom texte popíšeme. Na základe štatistickej teórie učenia, majú SVM tri základné výhody oproti sieťam používajúce metódu spätného šírenia chýb (ktoré sa najčastejšie využívajú).
Po prvé, majú väčšiu generalizačnú (zovšeobecnovaciu) schopnosť, inými slovami dokážu vyprodukovať oveľa presnejšie funčné hodnoty, predpovede či klasifikácie pre údaje, ktoré sa nepodieľali na učení siete. Po druhé, ich architektúra a váhy majú zaručenú unikátnosť a celkovú optimálnosť (odpadá pri nich problém lokálnych miním). A nakoniec, ich trénovanie je rýchlejšie. Z toho vyplýva, že navrhované SVM siete sú pre mnohé problémy robustnejšie, výkonnejšie, presnejšie a efektívnejšie ako existujúce MLP siete. Pre tieto vlastnosti ich možno odporučiť ako alternatívnu metódu, k už existujúcim modelom, ktoré sú určené na klasifikačné, regresné či prognózne úlohy v hydrológii a vo vodnom hospodárstve.
SVM využíva princíp tzv. štrukturálnej minimalizácie namiesto empirickej minimalizácie chyby – (Vapnik 1995) Pri trénovaní siete MLP, je jediným cieľom minimalizovať celkovú empirickú chybu. Pri SVM sa simultánne minimalizuje táto chyba aj zložitosť modelu. Použitie tohto princípu vedie k zostaveniu presnejších predpovedí pre dáta, ktoré neboli použité pri trénovaní SVM. Ďalším dôležitým rozdielom je spôsob určenia architektúry modelu a jeho váh. Architektúru MLP siete a jej váhy sa určujú na základe postupu pokus – omyl, čiže iteratívneho procesu a metódy spätného šírenia chýb, čo je časovo náročné. Vápnik (1995) upustil od časovo náročných trénovacích procesov a vyjadril určenie architektúry a váh ako kvadratický optimalizačný problém, ktorý možno riešiť štandardnými deterministickými algoritmami. A čo viac, pri nelineárnych úlohách SVM mapuje vstupný priestor do viacrozmerného funkčného priestoru (zobrazením vstupu do dostatočne veľkého funkčného priestoru sa klasifikačný problém stane lineárne separovateľný – Obr. 1.) a potom použije jadrovú (kernel)

funkciu namiesto viacrozmerného skalárneho súčinu. To má za následok, že výsledné riešenie je jednoduchšie, unikátne a optimálne bez problémov s lokálnym minimom.
Neurónové siete sa obvykle trénujú na všetkých dostupných dátach. SVM začínajú na všetkých dátach, ale blízko konca trénovania použijú iba malú podmnožinu najvhodnejších vektorov blízko okrajov rozhodovania (tzv. podporných vektorov). Hranica „užitočnosti" pre podporné vektory je dynamicky nastavovaná v priebehu učenia. Výhodou takéhoto prístupu je rýchlejšie učenie, vyššia presnosť konečných riešení a malé množstvo podporných vektorov, ktoré ovplyvňujú riešenie či už klasifikačného alebo regresného problému. Tento prístup prispieva ku zvýšeniu rýchlosti učenia a tým aj zníženiu doby behu učiaceho cyklu, ale tiež k vyššej presnosti.



Obr. 1. Príklad mapovania vstupných dát z dvoj do viacrozmerného priestoru
a) zložitejšie - nelineárne rozdelenie v dvojrozmernom priestore, b) lineárne rozdelenie tried vo viacrozmernom priestore (pre prípad klasifikácie)
Fig. 1. Example of mapping input data from two to multi-dimensional space
a) complex - non-linear distribution in two-dimensional space, b)linear division of classes in multidimensional space (in case of classification)

Predchádzajúci text a obrázok 1 vysvetľoval základné princípy SVM na klasifikačnom probléme, čo je ich prvotný spôsob aplikácie, v ďalšom stručne uvedieme použitie SVM pri regresnom probléme. Podrobný popis uvádza Vápnik (1997).
Vstupné údaje zadané do modelu SVM možno v zjednodušenom tvare zapísať ako množinu dátových bodov $G = \{(x_i, d_i)\}_i^n$, kde x_i je vstupný vektor, d_i je požadovaná hodnota a n je celkový počet dát.
SVM aproximuje túto funkciu nasledovne:

$$y = f(x) = w\ \varnothing(x) + b \qquad (3)$$

Čo predstavuje hľadanú hyper-rovinu vo vysokorozmernom priestore. Člen $\varnothing(x)$ je viacdimenzionálny

funkčný priestor, ktorý je nelineárne mapovaný zo vstupného priestoru $\boldsymbol{X}$. Koeficienty w a b, ktoré sú

v konečnom dôsledku analogické váhovým parametrom MLP sa stanovia minimalizovaním nasledovného výrazu:

$$R_{SVM}(C) = C\frac{1}{n}\sum_{i=1}^{n} L(d_i, y_i) + \frac{1}{2}\|w\|^2 \quad (4)$$

Za podmienky:

$L\varepsilon(d,y) = |d-1| - \varepsilon|d-1| \geq \varepsilon$ alebo je rovná 0

V regulovanej funkcii rizika z rovnice číslo 4 je prvá časť $C\frac{1}{n}\sum_{i=1}^{n} L(d_i, y_i)$ empirická chyba, meraná funkciou Lε. Druhá časť $0.5\|w\|^2$ je regulačný člen, ktorý je odvodený z princípu štrukturálnej minimalizácie. Regulačná konštanta *C* určuje kompromis medzi empirickým rizikom a regulačným členom. Zvyšovanie hodnoty *C* bude viesť ku relatívnemu zvyšovaniu významnosti empirickej chyby vzhľadom na regulačný člen ovplyvňujúci zložitosť modelu. Parameter *ε* je ekvivalentný presnosti aproximácie trénovacích dátových bodov.

Na získanie hodnôt hľadaných parametrov *w* a *b*, je rovnica 4, transformovaná na funkciu, vyjadrenú rovnicou číslo 5 zavedením pozitívnych druhotných premenných ζ_i a ζ_i^*.

$$R_{SVM} = (w, \zeta^{(*)}) = \frac{1}{2}\|w\|^2 + C\sum_{i=1}^{n}(\zeta_i + \zeta_i^*) \quad (5)$$

za podmienky, že : $d_i - w\phi(x_i) - b_i \leq \varepsilon + \zeta_i$

$$w\phi(x_i) + b_i - d_i \leq \varepsilon + \zeta_i^*, \ \zeta_i^* \geq 0$$

Napokon, zavedením Lagrangeových multiplikátorov a využitím obmedzení optimalneho problému sa rozhodovacia funkcia z rovnice číslo 3 zmení na nasledovnú explicitnú formu:

$$f(x, a_i, a_i^*) = \sum_{i=1}^{n}(a_i - a_i^*)K(x, x_i) + b \quad (6)$$

Funkcia $K(x_i,x_j)$ je definovaná ako jadrová funkcia (Kernel function). Hodnota jadrovej funkcie je ekvivalentná skalárnemu súčinu dvoch vektorov X_i a X_j vo funkčnom priestore $\varphi(x_i)$, $\varphi(x_j)$ a teda $K(x_i,x_j) = \varphi(x_i) * \varphi(x_j)$. Výhodou použitia jadrovej funkcie je že ľubovoľne veľký vstupný priestor, môže byť vypočítaný bez toho aby sme museli jeho mapovanie $\varphi(x)$ počítať explicitne. Každá funkcia zodpovedajúca Mercerovým podmienkam (Mukherjee et al. 1997) môže byť použitá ako jadrová funkcia. Typickým príkladom takýchto funkcií sú polynomický kernel $K(x,y) = (x^*y+1)^d$ alebo Gaussianovský kernel $K(x,y) = exp(-1/\delta^2(x-y)^2)$, kde *d* je stupeň polynómu jadra a δ^2 je šírka pásma Gaussianovho jadra.

Zmysel jadrových funkcií je teda v tom, že pre optimalizáciu možno použiť priamo kernelovú funkciu pričom nie je nutné poznať ani počítať transformáciu do viacrozmerného priestoru, čo je v literatúre uvádzané ako tzv. „kernel trick“. Týmto je umožnené výpočtovo zvládnuť nelineárnu klasifikáciu a regresiu pomocou SVM.

Pretože jadrový parameter implicitne definuje štruktúru viacrozmerného funkčného priestoru $\varphi(x)$ a určuje komplexnosť výsledného riešenia, musíme zvoliť vhodný typ (čo je spolu s *C* a *ε* jeden z volených parametrov výpočtu pomocou SVM). Vo tejto práci bola ako kernelová (jadrová) funkcia použitá RBF (Radial Basis Function), ktorá dokáže na rozdiel od lineárnych kernelových funkcií zvládnuť aj prípady, kedy je vzťah medzi triedami dát nelineárny. Kernelová funkcia RBF je definovaná ako:

$$K(x_i; x_j) = \exp\left(\gamma\|x_i - x_j\|^2\right)$$

kde *γ* je parameter kernelovej funkcie RBF určujúci transformáciu.

Určenie bodov vlhkostných retenčných kriviek pedotransferovými funkciami

Pri tvorbe PTF pôd Záhorskej nížiny bola použitá multilineárna regresia. Zo vstupného trénovacieho súboru sme regresnou analýzou stanovili regresné koeficienty a tie sme potom použili pri výpočte s testovacími dátami nasledovne:

$$\theta_{hw} = A * I.kat. + B * II.kat + C * III.kat + D * IV.kat + E * \rho_d + F \qquad [cm^3.cm^{-3}] \qquad (7)$$

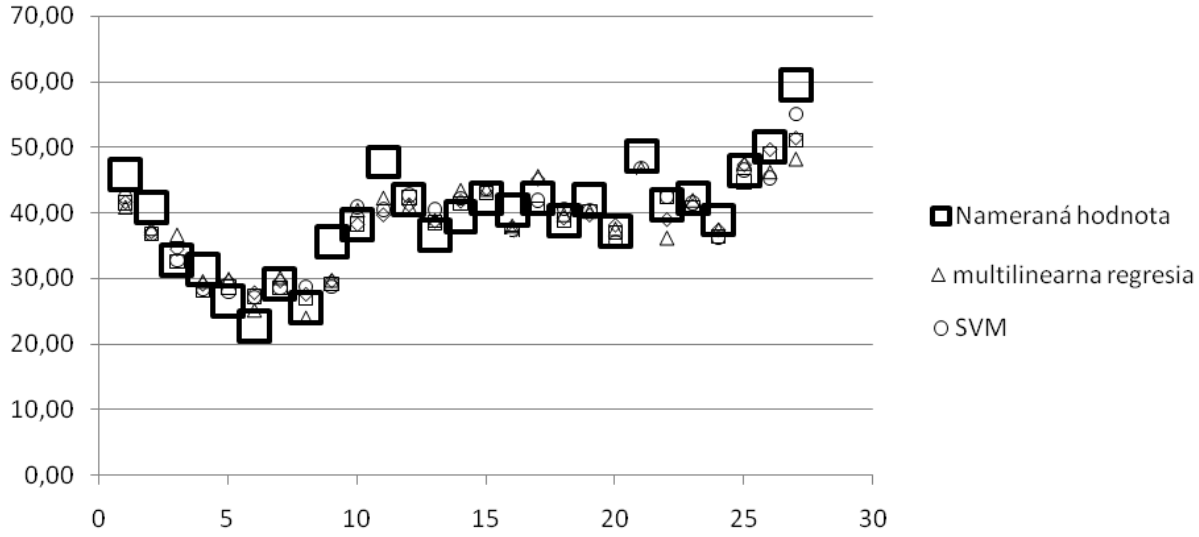
kde: θ_{hw} – vlhkosť [$cm^3.cm^{-3}$] pri konkrétnej hodnote vlhkostného potenciálu h_w [cm]
I.kat. – % obsah ílovitých častíc (d < 0.01 mm),
II.kat. – % obsah prachových častíc (0,01 – 0.05 mm),
III.kat. – % obsah častíc práškového piesku (0.05 – 0.1mm),
IV.kat. – % obsah piesku (0.1 – 2.0 mm),
ρ_d – redukovaná objemová hmotnosť [$g.cm^{-3}$],
A, B, C, D, E,F – koeficienty zistené regresnou analýzou.

Pedotransferové funkcie boli získané pre objemové vlhkosti θ_{hw} (závisle premenné) vo vlhkostných potenciáloch h_w = -2.5, -56, -209, -558, -976, -3060, -15300 cm v závislosti od percentuálneho obsahu I., II., III. a IV. zrnitostnej kategórie podľa Kopeckého a od objemovej hmotnosti redukovanej ρ_d (čo je 5 nezávislých premenných).

Výpočet VRK popísanými metódami

Na testovanie a porovnanie rôznych metód stanovenia VRK bol použitý súbor dát o veľkosti 133 vzoriek o 12 parametroch. Údaje sú rozdelené na trénovacie (106x12) a testovacie (27x12) dáta a na vstupné a požadované (výsledné) dáta. Tento krok je veľmi dôležitý a využijú sa pri ňom bežné štatistické postupy za účelom poznania charakteristických vlastností dostupných údajov. Medzi vstupné údaje patria zrnitostné kategórie I až IV triedy a redukovaná objemová hmotnosť (ρ_d). Ako výstup bola určená vlhkosť pre konkrétne vlhkostné potenciály h_w = -2.5, -56, -209, -558, -3060 a -15300 cm. Výpočet sme najskôr realizovali pomocou SVM.

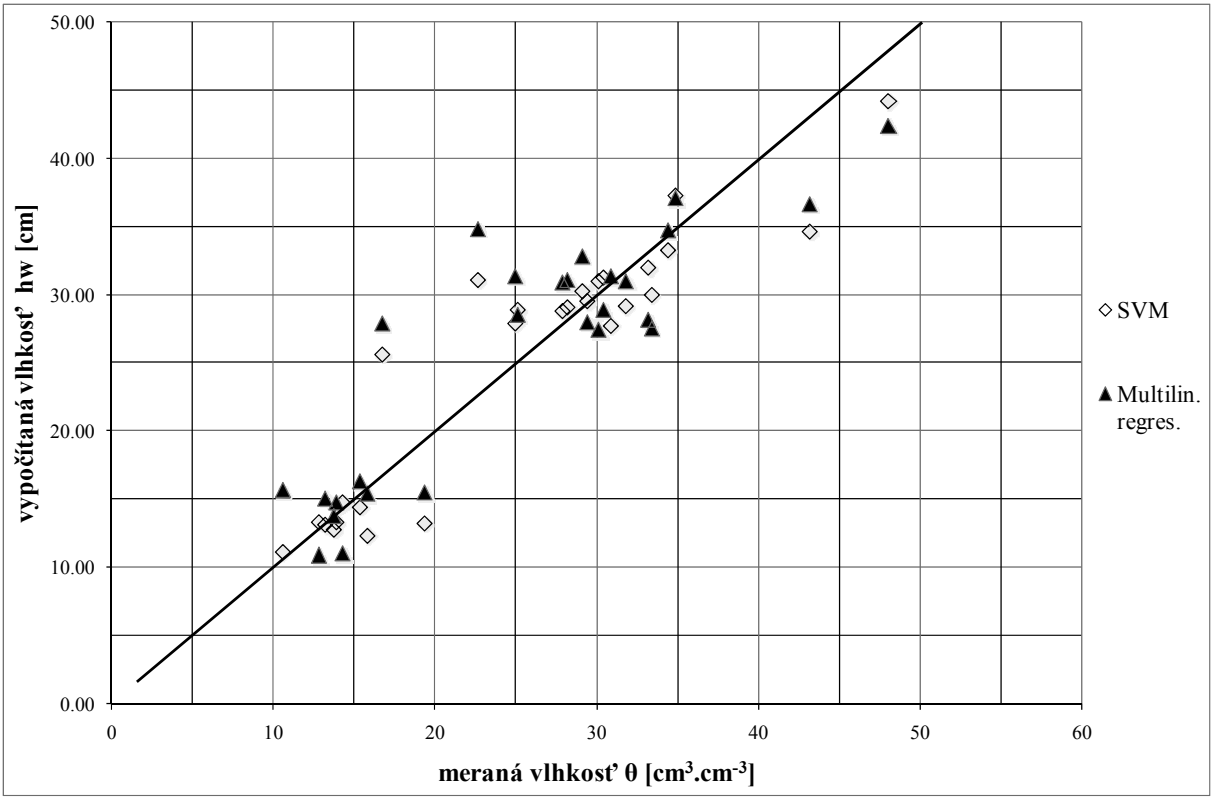
Prvým krokom bol návrh architektúry SVM, čiže zadefinovanie riešeného problému. V našom prípade regresiu s použitím učiacej metódy Epsilon SVR. Výpočet pomocou SVM bol prevedený pomocou príslušného algoritmu, popísaného v predchádzajúcej časti. Samotná sieť mala okrem preddefinovaných komponentov aj niekoľko nastaviteľných parametrov ako typ kernel (jadrovej) funkcie - radial basis function, počet jej stupňov 3, parameter epsilon 0.1 a koeficient penalizácie 2 (Penality cost). Optimálne nastavenie jednotlivých parametrov sme stanovili na základe metódy pokus – omyl. Ako konkrétna implementácia bola využitá tzv. LIBSVM, ktorá je bližšie popísaná v (Chang, Lin, 2001).
Potom sa sieť natrénovala s trénovacími dátami. Po natrénovaní siete, potrebné posúdiť kvalitu výsledkov na dátovom súbore, ktorý neprispel k výstavbe modelu. Na tieto účely boli k dispozícii tie isté testovacie dáta – 27 údajov, ktoré porovnáme s dátami vypočítanými pomocou SVM. Pre overenie správnosti výsledkov SVM vykonané aj výpočty pomocou multilineárnej regresie. Názorné grafické porovnanie výsledkov jednotlivých výpočtových metód je zobrazené na obr. 3 a je uvedené ako príklad pre vlhkostný potenciál h_w = -976 cm. Hodnoty korelačných koeficientov medzi nameranými a vypočítanými údajmi sú v tabuľke 1. Korelačný koeficient, alebo koeficient korelácie, je miera intenzity (teda tesnosti) lineárnej korelácie (pri priamkovej regresii) medzi dvoma náhodnými veličinami, ak. Môže nadobúdať hodnoty z intervalu . Ak je korelačný koeficient rovný 0, znamená to, že medzi skúmanými veličinami nie je žiadny vzťah. Ak je rovný jednej, tak sú veličiny navzájom priamo závislé, nárast jednej veličiny spôsobuje nárast druhej. V prípade, že je rovný −1, nárast jednej veličiny spôsobuje pokles druhej veličiny. Korelácia je vysoká, keď sa dá meranými bodmi v rovine metódou najmenších štvorcov "dobre preložiť" priamka.



Obr. 2. Porovnanie nameraných a vypočítaných hodnôt pomocou SVM a multilin. regresie na testovacom súbore
Fig. 2. Comparison of measured and calculated values using SVM and multilin. regresion on the test set

Tabuľka 1. Korelačné koeficienty nameraných a vypočítaných údajov
Tab. 1. Correlation coefficients measured and calculated data

korelácia pre h_w	**SVM**	**Multilinear. reg.**
-2.5	0,92305	0,89808
-56	0,93301	0,91915
-209	0,93651	0,90943
-558	0,96207	0,92931
-976	0,92695	0,88281
-3060	0,93096	0,90545
-15300	0,91335	0,88726
Σ	6,53	6,33



Obr. 3. Porovnanie nameraných a vypočítaných hodnôt pre vlhkostný potenciál h_w = - 976cm
Fig. 3. Comparison of measured and calculated values for moisture potential h_w = - 976cm

Záver

Hlavnou výhodou UNS je ich schopnosť učiť sa zo vzorových údajov. Pomocou UNS je možné popísať aj zložité vzťahy pre vysokodimenzionálne dáta, ktoré nie je možné popísať jednoduchými pravidlami alebo zákonitosti platiace v množine týchto dát nie je ľahké objaviť napr. štatistickými metódami.

V tejto práci sme stanovovali body pedotransférnej funkcie dvoma metódami a to pomocou mechanizmu podporných vektorov (SVM – support vector machines) a multilineárnou regresiou, čo sú možné alternatívy k náročným laboratórnym meraniam. Na porovnanie výsledkov a zhody medzi nameranými a modelovanými dátami bola vykonaná analýza pomocou korelačného koeficientu. Ten bol pre všetky namerané vlhkosti pre konkrétne vlhkostné potenciály (h_w = -2.5, -56, -209, -558, -3060 a -15300 cm) u SVM najpresnejší, v porovnaní s multilneárnou regresiou. Celková suma koeficientov jednotlivých metód dokazuje, že SVM si viedli v našom pokuse najlepšie s výsledkom súčtu korelačných koeficientov 6,53 oproti multilineárnej regresii - 6,33. To naznačuje ich lepšiu presnosť pri výpočtoch aj s malým množstvom vstupných dát ako aj ich väčšiu výpočtovú stabilitu a rýchlosť výpočtu. Z týchto výsledkov ako aj z grafických vyhodnotení vyplýva, že učiaca technika SVM je v porovnaní s multilineárnou regresiou presnejšia. Ukázala sa ako vhodná alternatívna metóda pri vodohospodárskych výpočtoch a predpovediach, avšak bolo by potrebné tento výsledok overiť aj v iných lokalitách a pre iné problémy.

Poďakovanie

Tento príspevok bol vypracovaný na základe podpory z projektu APVV-0271-07, APVV- 0443-07, a LPP-0319-09

Literatúra

Šimor, V., Čistý, M., Hlavčová, K. (2009): Predpoveď prietokov na povodí rieky Hron pomocou umelých neurónových sietí. In: Acta hydrologica slovaca. - ISSN 1335-6291. - Roč.10, č. 1, s. 159 - 170
Brooks, R. H., Corey, A., T. (1964): Hydraulic properties of porous media. Hydrol. Pap.3. Colorado State Univ., Fort Collins.
Gupta, S. C., Larson, W. E. (1979): Estimating soil water characteristic from particle size distribution, organoc mater percent, and bulk density. Water Resour., 15, 1633-1635.
Rawls, W. J., Brakensiek, D. L., Saxton, K. E. (1982): Estimation of soil water properties, Trans.ASAE, 108, 1316-1320.
Renger, M. (1971): The estimation of pore size distribution from texture, organic matter content and bulk density, Z. Kluturtech. Flubereinig, 130, 53-67.
Wösten, J. H., Lilly, A., Nemes, A., Bas, C. (1998): Using existing soil date to derive hydrauloc parameters for simulation models enviromental studies and land use planning, DLO-starimg centre, Wageningen, 103 s.
Bouma, J., Van Lanen, j. A. J. (1987): Treánsfer function and treshhold values: from soil characteristics to land qualities.
Skalová, J. (2001): Pedotransferové funkcie pôd Záhorskej nížiny a ich aplikácia pri modelovaní vodného režimu pôd, Svf STU Bratislava, 112 s.
Skalová, J., Branislav, J. (2008): Soil water assessment in Morava basin. In: Cereal Research Communications. - ISSN 0133-3720. - Vol. 36, Supplementum, Part I, s. 243-246.
Skalová, J., Branislav J. (2007): Soil water regime evaluation. In: Pollution and water resources: Columbia University seminar proceedings.Vol.XXXVII.2006-2007. - Columbia: Columbia University Seminar, ISBN 978-80-89139-12-5. - S. 135-142.
Velebný, V., Skalová, J. (1994): Pôdohospodárstvo a pedológia, Bratislava.
Skalová, J. (2003): Použitie pedotransferových funkcií pri diagnostike vodného režimu pôd. Acta Hydrologica Slovaca, UH SAV Bratislava, ISSN 1335-6291, ročník 4, č.2, s. 300–306.
Kvasnička, V., Beňušková, Ľ., Pospíchal, J.,Farkaš, I., Tiňo, P., Kráľ, A. (1997): Úvod do teórie neurónových sietí. Iris, Bratislava, ISBN 80-88778-30-1.
Vladimir N. Vapnik (1995): The Nature of Statistical Learning Theory, Springer.
C.-C.Chang and C.-J. Lin. (2001): LIBSVM: a library for support vector machines.
http://www.csie.ntu.edu.tw/~cjlin
*http://*www.root.cz/clanky/neuronove-siete-su-ciernou-skrinkou/

Alternative methods for determining water content in soil

Hydrological processes are generally nonlinear and are governed by general physical causalities. Their variability covers a number of spatial and temporal levels. That's the reason why it is difficult to observe, describe and model them. Nowadays one of the best ways of examining these processes in large areas possible is through modeling.
The advantage of modeling using artificial neural networks is their ability to learn from the model data and the ability to generalize knowledge from them, often without detailed knowledge of various state variables of the process, makes them suitable alternative tool to address complex processes. Information's about soil water content are in the planning of water resources and management very valuable. Modeling and predicting soil water transfer is very important in agriculture or hydrology. One of the most important hydro physical characteristics of soil which is entering hydraulic and hydrological models is water retention curve (VRK), which reflects the energy dependence of soil water from the water content, thus the relationship between soil moisture and moisture potential. The method of determining the water retention curve points in laboratory conditions is difficult for experimental equipment and also time required for the use of this method is very laborious. In soil

physics, therefore, developed the method for determining soil hydro physical characteristics depending on its granularity composition, organic matter content and bulk density. For these models (or relations) have been coined a title pedotransfer function (PTF), these functions indicate different soil characteristics and properties to relationship from one to another. The submitted work compares the creation of functional dependencies using support vector machines (SVM) and standard multi-linear regression. PTF Záhorská lowland soils and their application to modeling soil water regime developed Skalová (2001) where we assume the input data and the percentage of granularity categories I to IV according to Kopecky method, reduced volume weight (ρd) and humidity potentials hw = -2.5; -56, -209, -558, -976, -3060, -15300 cm specified in laboratory for determination and testing the regression dependency.

To compare the results between measured and modeled data, we create an analysis using correlation coefficient. Coefficient was for all the measured moisture content for concrete moisture potentials most accurate by using SVM, compared with multi-linear regression. The total coefficients of these methods shows that the SVM in our experiment resulted from all methods with the best result of the sum of the correlation coefficients of 6.53 and multilinear regression - 6.33. This suggests the greater accuracy of the calculations with little input data as well as their greater stability and speed of computer calculation. From these results, as well as the graphical evaluation shows that the learning technique SVM is more accurate than multilinear regression. Therefore SVM has proved as suitable alternative method for hydraulic calculations and predictions, but it would be necessary to verify this result in other locations and other problems.

Ing. Martin Suchár
Katedra vodného hospodárstva krajiny
Stavebná fakulta Slovenskej technickej univerzity v Bratislave
Radlinského 11
813 68 Bratislava
E-mail: martin.suchar@stuba.sk